Helsinki
Center
of
Economic
Research

Discussion Papers

Residual-based diagnostic tests for noninvertible ARMA models

Juho Nyholm
University of Helsinki and HECER

Discussion Paper No. 416
August 2017

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND,
Tel +358-2941-28780, E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 416

Residual-based diagnostic tests for noninvertible ARMA models*

Abstract

This paper proposes two residual-based diagnostic tests for noninvertible ARMA models. The tests are analogous to the portmanteau tests developed by Box and Pierce (1970), Ljung and Box (1978) and McLeod and Li (1983) in the conventional invertible case.
We derive the asymptotic chi-squared distributions for the tests and study the size and power properties in a Monte Carlo simulation study.
An empirical application employing financial time series data points out the usefulness of noninvertible ARMA model in analyzing stock returns and the use of the proposed test statistics.

JEL Classification: C22, C52

Keywords: Non-Gaussian time series, noninvertible ARMA model, model selection

Juho Nyholm

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: juho.nyholm@helsinki.fi

* I am grateful for the guidance and suggestions of Markku Lanne, Mika Meitz and Pentti Saikkonen. I am also thankful to Tim Bollerslev and Henri Nyberg, and all the participants at Financial Time Series Econometrics lunch group at Duke University, HECER Econometrics Workshops, 9th Nordic Econometric Meeting, and SoFiE Financial Econometrics summer school for useful comments. Financial support from Academy of Finland, Yrjö Jahnsson Foundation and OP-Pohjola Group Research Foundation is gratefully acknowledged.

1 Introduction

The usefulness of noninvertible time series models in economic research has first been pointed out in the seminal research by Hansen and Sargent (1981) and Hansen and Sargent (1991) where the markets are modeled under the assumption that the agents know more than the modeling econometrician. More recently the nonfundamentalness issue has risen in asset pricing models (Kasa, Walker, and Whiteman, 2014), fiscal foresight models (Leeper, Walker, and Yang, 2013), news shocks models (Blanchard and Perotti, 2002; Forni and Gambetti, 2014), and in permanent income models (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007). It is common for all these models that they have a noninvertible linearised solution. A comprehensive survey on this topic in economic theory can be found in Alessi, Barigozzi, and Capasso (2011).

Empirical evidence of the good fit of noncausal time series models to economic time series can be found, for example, in Gouriéroux and Zakoïan (2013), Lof (2013) and Nyberg, Lanne, and Saarinen (2012), and for the noninvertible case, in Andrews, Calder, and Davis (2009), Breidt, Davis, and Trindade (2001), and Huang and Pawitan (2000). Lanne, Meitz, and Saikkonen (2013) point out that noninvertible models are potentially capable of capturing the nonlinearities in the time series as they are driven by the iid error terms in nonlinear manner. For example, these models are shown to control for mild heteroskedasticity commonly encountered in financial time series. They are also capable of producing time series that are at most very mildly autocorrelated but still dependent in a nonlinear way. These nonlinearities can not be controlled by the conventional causal and invertible models, as the lack of autocorrelation automatically implies independence of the observations.

Although the evidence for usefulness of noninvertible and noncausal models is growing, the issue of model selection and model adequacy testing has been addressed surprisingly little in the literature. For noninvertible models, this paper is, to the best of our knowledge, the first attempt to provide asymptotically valid checks for the adequacy of the selected model. For noncausal model, see Cui, Fisher, and Wu (2014) for related results.

After the seminal work of Box and Pierce (1970), Ljung and Box (1978), and McLeod and Li (1983), the most common practice in model evaluation has been to compare the sample autocorrelation functions of the model residuals and squared residuals to the appropriate asymptotic confidence bands under the assumption that the correct model has been selected. A statistically significant departure from zero of the autocorrelation function indicates that there is still information in the residuals that could be used in modeling. Thus, it indicates that the selected model should be changed. However, the asymptotic properties of these tests have not been studied under the assumption of non-invertibility.

In this article we derive the asymptotic distributions of the sample autocorrelation function of the residuals and the squared residuals, obtained by maximum likelihood estimation of a causal and noninvertible ARMA model. That is, we assume that the roots of the AR polynomial lie outside the unit circle and the roots of the MA polynomial are located inside the unit circle. Using this asymptotic result, the main contributions are χ^2-distributed test statistics to detect dependencies in the residuals with different lag lengths.

We say that a test statistic is invariant to estimation uncertainty if the distribution of the error terms is the only driver of the asymptotic properties of it. Vice versa, a test statistic is not invariant to estimation uncertainty if the asymptotic properties of the estimators affect the asymptotic properties of the test. We show that the test for autocorrelation in the residuals is not invariant to the estimation uncertainty, whereas the test for autocorrelation in the squared residuals is. That is, the properties of the McLeod - Li test are unchanged, but the autocorrelation test needs to be adjusted to accommodate the properties of the parameter estimates of noninvertible model.

As an empirical application, we build on the work by Lanne et al. (2013). They propose a two step procedure for testing the predictability in several financial time series. As a first step, they test if the data is white noise against the alternative of correlated data. If this hypothesis is not rejected, it is possible to test a stronger hypothesis of independence against the white noise hypothesis. They provide illustrative and descriptive evidence that the ARMA(1,1) model is adequate for the series considered, and that it is sensible to base their predictability testing procedure on this model. Using the formal tests of the present paper, we can conclude that indeed, a noninvertible ARMA(1,1) model seems to be a good description of the data generating process for most of the financial time series considered by Lanne et al. (2013).

The rest of the paper is organized as follows. Section 2 describes the noninvertible ARMA model in detail and discusses briefly its maximum likelihood estimation. Section 3 introduces the test statistics considered and derives their asymptotic properties. In Section 4 we conduct a Monte Carlo simulation experiment to study the small sample properties of the proposed test statistics. Section 5 provides an empirical example using financial time series data and Section 6 concludes. High level assumption are left for the appendices to ease the reading, as well as some intermediate results and lemmas used in the proof of the main theorem. These results and lemmas are proved in a supplementary appendix that is available upon request.

A few notational conventions are given. Convergence in probability and in distribution are denoted by "$\xrightarrow{p}$" and "$\xrightarrow{d}$", respectively. All vectors are column vectors unless otherwise indicated. That is, $x = (x_1, \ldots, x_h)$ is a column vector consisting of h elements that are either scalars or column vectors. The L_r-norm is denoted by $\|\cdot\|_r$. For a random variable x, $\|x\|_r = \mathrm{E}\left[|x|^r\right]^{1/r}$. Abbreviation *a.s.* stands for *almost sure*. Identity matrix of size m is denoted by $I_{m\times m}$ and a matrix of size $k \times l$ where all the elements are zeros is denoted by $0_{k\times l}$.

2 The Noninvertible ARMA Model

Maximum likelihood estimation of noninvertible ARMA models has been discussed, among others, by Lii and Rosenblatt (1996), Rosenblatt (2012), and Meitz and Saikkonen (2013). We study the residuals of an estimated noninvertible ARMA(P, Q) process[1]

$$a_0(B)y_t = b_0(B^{-1})\varepsilon_t, \tag{1}$$

where B and B^{-1} denote the backward shift and forward shift operators, $a_0(z) = 1 - a_{0,1}z - \cdots - a_{0,P}z^P$ is an autoregressive (AR) polynomial of order P and has its roots outside the unit circle, $b_0(z) = 1 - b_{0,1}z - \cdots - b_{0,Q}z^Q$ is a moving average (MA) polynomial of order Q with all of its roots inside the unit circle, and $\varepsilon_t = \sigma\eta_t$ is a non-Gaussian iid error term process with $0 < \sigma^2 < \infty$ and η_t having mean zero and variance one. It is assumed that η_t has a symmetric distribution function $f_\eta(x; \lambda_0)$, where λ_0 is a

[1]Throughout the paper we use sub-index zero to distinguish the true but unknown parameter value from other possible parameter values.

$d \times 1$ parameter vector. The Gaussian distribution must be ruled out because for each noninvertible model, there is always an invertible model with exactly the same second order properties and in the Gaussian case these models cannot be distinguished from each other (for a thorough discussion, see Rosenblatt (2012), Chapter 2).

Note that in (1), the MA polynomial is defined in terms of z^{-1} instead of z. We follow the example of Meitz and Saikkonen (2013), and write the model in the form where its dependence on the future error terms is displayed explicitly. Model (1) has an MA(∞) representation in terms of Q future, the present, and the infinite history of the error terms ε_t. It also has an AR(∞) representation in terms of P lagged, the present, and the infinite future of the output process y_t,

$$y_t = \sum_{j=-Q}^{\infty} \psi_{0,j}\varepsilon_{t-j} \quad \text{and} \quad \varepsilon_t = \sum_{j=-P}^{\infty} \pi_{0,j} y_{t+j}. \tag{2}$$

The coefficients $\psi_{0,j}$ and $\pi_{0,j}$ are geometrically decaying coefficients of the Laurent series expansions of $a_0(z)^{-1}b_0(z^{-1})$ and $a_0(z)b_0(z^{-1})^{-1}$, respectively.[2]

Let us define a counterpart of the process ε_t defined for all the parameter values $\theta = (a_1, \ldots, a_P, b_1, \ldots, b_Q, \sigma, \lambda) \in \Theta$, where Θ is the permissible parameter space defined in Assumption 2 in Appendix A. In analogue to (2), set[3]

$$u_t(\theta) = \frac{a(B)}{b(B^{-1})} y_t = \sum_{j=-P}^{\infty} \pi_j y_{t+j}, \tag{3}$$

where $a(z) = 1 - a_1 z - \cdots - a_P z^P$ and $b(z) = 1 - b_1 z - \cdots - b_Q z^Q$. This sum is well defined and the sequence of coefficients π_j decays geometrically.

Because the infinite past and future of the process y_t are not observable at time t, our diagnostic tests must be based on a *feasible* counterpart of the sequence $u_t(\theta)$, say $\tilde{u}_t(\theta)$. Let us assume that we have observed $\{y_t\}_{t=1-P}^{T}$. The feasible sequence $\tilde{u}_t(\theta)$ of size T is obtained by initializing by $\tilde{u}_{T+1}(\theta) = \cdots = \tilde{u}_{T+Q}(\theta) = 0$, and then solving top-down, for $t = T, \ldots, 1$,

$$\tilde{u}_t(\theta) = y_t - a_1 y_{t-1} - \cdots - a_P y_{t-P} + b_1 \tilde{u}_{t+1}(\theta) + \cdots + b_{t+Q} \tilde{u}_{t+Q}(\theta). \tag{4}$$

[2]See the end of Appendix D and especially Lemmas A.1. and A.2. in Meitz and Saikkonen (2013) for a throughout discussion of these series presentations.

[3]Polynomials $a(B)$ and $b(B^{-1})$ are defined by the parameter vector $\theta \neq \theta_0$ as $a(B) = 1 - a_1 B - \cdots - a_P B^P$ and $b(B^{-1}) = 1 - b_1 B^{-1} - \cdots - b_Q B^{-Q}$.

Regarding parameter estimation, Meitz and Saikkonen (2013) discuss maximum likelihood estimation of Model (1) with an error term assumed to follow an ARCH process. Our model is a simplified version of theirs and the asymptotic properties of the ML estimator are obtained in a very similar fashion as in their paper. These properties are listed in Proposition 1 below. Let $L_T(\theta)$ denote an approximation of the log-likelihood function of the model (for details, see Meitz and Saikkonen (2013)),

$$L_T(\theta) = T^{-1}\sum_{t=1}^{T} l_t(\theta) \quad \text{with} \quad l_t(\theta) = \log f_\eta\left(\sigma^{-1}u_t(\theta);\lambda\right) - \log\sigma, \tag{5}$$

and let $L_{\theta,T}(\theta)$ and $L_{\theta\theta,T}(\theta)$ denote the first and second order derivatives of the log-likelihood with respect to the parameter vector $\theta = (a_1,\ldots,a_P,b_1,\ldots,b_Q,\sigma,\lambda)$. Estimation and statistical inference is based on feasible versions of these quantities denoted by $\tilde{L}_T(\theta)$, $\tilde{L}_{\theta,T}(\theta)$ and $\tilde{L}_{\theta\theta,T}(\theta)$, which are obtained by replacing $u_t(\theta)$ by its feasible counterpart $\tilde{u}_t(\theta)$ in the log-likelihood function (exact expressions for these quantities can be found in Appendix C).

The following proposition contains conventional properties of a (local) maximum likelihood estimator. We omit the proof for brevity, but the arguments are very similar to those used by Meitz and Saikkonen (2013).

Proposition 1. *Under Assumptions 2 and 3 in Appendix A,*

1. $\lim_{T\to\infty} Cov\left(T^{1/2}L_{\theta,T}(\theta_0)\right) = \ell(\theta_0)$ *where* $\ell(\theta_0)$ *is positive definite and* $\ell(\theta_0) = -E[l_{\theta\theta,t}(\theta_0)]$,
2. $T^{1/2}L_{\theta,T}(\theta_0) \xrightarrow{d} N(0,\ell(\theta_0))$,
3. $\sup_{\theta\in\Theta_0}|L_{\theta\theta,T}(\theta) - \mathcal{J}(\theta)| \to 0$ *a.s. as* $T\to\infty$, *where* $\mathcal{J}(\theta) = E[l_{\theta\theta,t}(\theta)]$ *is finite and continuous at* θ_0, *and* Θ_0 *(defined in Assumption 2 in Appendix A) is some compact and convex set containing* θ_0,
4. *there exists a sequence of solutions* $\tilde{\theta}_T$ *to the likelihood equations* $\tilde{L}_{\theta,T}(\theta) = 0$ *s.t.* $T^{1/2}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0,\ell(\theta_0)^{-1})$, *and*
5. *there is a consistent estimator for the asymptotic covariance matrix given by the inverse of the Hessian,* $-\tilde{L}_{\theta\theta,T}(\tilde{\theta}_T)^{-1} \to \ell(\theta_0)^{-1}$ *a.s. as* $T\to\infty$.

These asymptotic properties of the ML estimator are the main ingredients for the asymptotic behavior of the test statistics we derive in the next section.

3 Diagnostic Tests

Consider the following expressions related to the autocorrelation functions of the residuals and the squared residuals:[4]

$$\tilde{g}_{ac,t}(\tilde{\theta}_T) \stackrel{def}{=} \tilde{u}_t(\tilde{\theta}_T)\begin{bmatrix}\tilde{u}_{t-1}(\tilde{\theta}_T)\\ \vdots \\ \tilde{u}_{t-m}(\tilde{\theta}_T)\end{bmatrix} \quad \text{and} \quad \tilde{g}_{hs,t}(\tilde{\theta}_T) \stackrel{def}{=} \left(\tilde{u}_t(\tilde{\theta}_T)^2 - \tilde{\sigma}_T^2\right)\begin{bmatrix}\tilde{u}_{t-1}(\tilde{\theta}_T)^2 - \tilde{\sigma}_T^2\\ \vdots \\ \tilde{u}_{t-m}(\tilde{\theta}_T)^2 - \tilde{\sigma}_T^2\end{bmatrix} \tag{6}$$

For $i \in \{ac, hs\}$, define

$$\tilde{q}_{i,T}(\tilde{\theta}_T) \stackrel{def}{=} (T-m)^{-1/2}\sum_{t=m+1}^{T}\tilde{g}_{i,t}(\tilde{\theta}_T).$$

The two test statistics we propose and study in this paper are

$$\begin{aligned} Q_{ac,T} &\stackrel{def}{=} \tilde{q}_{ac,T}(\tilde{\theta}_T)'\tilde{\Omega}_{ac,T}^{-1}\tilde{q}_{ac,T}(\tilde{\theta}_T) \quad \text{and} \\ Q_{hs,T} &\stackrel{def}{=} \tilde{q}_{hs,T}(\tilde{\theta}_T)'\tilde{\Omega}_{hs,T}^{-1}\tilde{q}_{hs,T}(\tilde{\theta}_T), \end{aligned}$$

where the positive definite matrix $\tilde{\Omega}_{i,T}$ estimates consistently the asymptotic covariance of $\tilde{q}_{i,T}(\tilde{\theta}_T)$. The exact forms of these matrices are given in Appendix E. The main result of the paper (to be presented below) states that under H_0,

H_0 : The true data generating process is Model (1), it satisfies Assumptions 1 (a), 2, and 3 in Appendix A, and the estimated model is correctly specified,

the $Q_{ac,T}$ test statistic is asymptotically χ^2_m-distributed, where m is the number of lags included in the test statistic. If we include Assumption 1 (b) to H_0, then also the $Q_{hs,T}$ test statistic has the same asymptotic distribution.

Correctly specified estimated model means that the likelihood function is derived using the true probability distribution of the error term, and that the orders of the polynomials $a(z)$ and $b(z^{-1})$ are correct. We expect the test statistics to have power against wide variety of misspecified models. The main example is the selection of the orders of the AR

[4]For ease of reading, we will suppress the dependence on the lag length m of the quantities we are defining in this section. Here m can be any positive integer smaller than T, and in practice it should be considered to be moderate in size. The exact number does not affect the derivations.

and MA polynomials, which can be potentially detected by both of the test statistics. We will also provide evidence in favor of the tests to be able to detect nonlinearities in the data generating process. Namely, ARCH type models can be detected by the heteroskedasticity test. On the other hand, we do not expect that deviation from the distributional assumption of the error term would be detected by these test statistics. Tests for these kinds of deviations from the null hypothesis are left for future work.

The proposed diagnostic tests for model adequacy are based on the ideas of Box and Pierce (1970) and Ljung and Box (1978) for checking the residual series and the ideas of McLeod and Li (1983) for checking the series of squared residuals. In line with these papers, we consider the asymptotic properties of the sample autocorrelation functions of these series. The key idea in the references above is to show how the asymptotic distribution of certain sample functions needs to be adjusted for the second order uncertainty of the estimation of autocorrelation functions, arising from the uncertainty in parameter estimation. We show how the asymptotic distribution of certain autocorrelation functions needs to be adjusted in order to capture the dynamics of the residual functions that are, in general, not only dependent on the past innovations, but also on the future innovations (see (3)).

The following theorem contains the main result of this paper.

Theorem 1. *(i) Under Assumptions 1 (a), 2 and 3 in Appendix A, we have, under H_0*

$$Q_{ac,T} \xrightarrow{d} \chi^2_m,$$

where m is the dimension of the test statistic.

(ii) If, in addition, Assumption 1 (b) in Appendix A holds, then, under H_0 also

$$Q_{hs,T} \xrightarrow{d} \chi^2_m.$$

The joint distribution of the sample autocorrelation function and the score vector describes the behavior of the test statistics. It incorporates both the uncertainty due to the randomness of the error terms and due to the uncertainty in the parameter estimators. We also show that the difference between using the observable feasible quantities instead of the unobservable ones is asymptotically negligible.

Intermediate results for the proof are given in Appendix D. For $i \in \{ac, hs\}$, let $g_{i,t}(\theta)$ denote the unfeasible counterpart of $\tilde{g}_{i,t}(\theta)$ in (6),

$$g_{ac,t}(\theta) \stackrel{def}{=} u_t(\theta) \begin{bmatrix} u_{t-1}(\theta) \\ \vdots \\ u_{t-m}(\theta) \end{bmatrix} \quad \text{and} \quad g_{hs,t}(\theta) \stackrel{def}{=} (u_t(\theta)^2 - \sigma^2) \begin{bmatrix} u_{t-1}(\theta)^2 - \sigma^2 \\ \vdots \\ u_{t-m}(\theta)^2 - \sigma^2 \end{bmatrix}. \tag{7}$$

Denote $g_{i,\theta,t}(\theta) \stackrel{def}{=} \frac{\partial}{\partial \theta'} g_{i,t}(\theta)$ and $\mathcal{G}_i(\theta) \stackrel{def}{=} E[g_{i,\theta,t}(\theta)]$, and let $\xi_{i,t}(\theta) \stackrel{def}{=} (l_{\theta,t}(\theta), g_{i,t}(\theta))$.

Proof of Theorem 1. Using Proposition 2 in Appendix D, the sequence $T^{-1/2} \sum_{t=1}^{T} \xi_{i,t}(\theta_0)$ has an asymptotic normal distribution with a positive definite covariance matrix $\Upsilon_i(\theta_0)$.

Mean value expansion of the feasible score around the true parameter value θ_0 gives

$$\tilde{L}_{\theta,T}(\tilde{\theta}_T) = \tilde{L}_{\theta,T}(\theta_0) + \tilde{L}_{\theta\theta,T}(\bar{\theta}_T)(\tilde{\theta}_T - \theta_0),$$

where the left hand side is zero by definition and $\bar{\theta}_{T,j} = \alpha_j \theta_{0,j} + (1 - \alpha_j)\tilde{\theta}_{T,j}$ for some $\alpha_j \in (0, 1)$ for all elements $j = 1, \dots, P+Q+1+d$. Assuming T sufficiently large so that $\tilde{\theta}_T \in \Theta_0$, also $\bar{\theta}_T \in \Theta_0$ by convexity of this set. Using the fact that $\tilde{L}_{\theta\theta,T}(\bar{\theta}_T)$ is positive definite with probability approaching one, and that the l.h.s. is zero by definition, we re-organize terms to get

$$\begin{aligned} T^{1/2}(\tilde{\theta}_T - \theta_0) &= - \tilde{L}^{-1}_{\theta\theta,T}(\bar{\theta}_T) T^{1/2} \tilde{L}_{\theta,T}(\theta_0) \\ &= - L^{-1}_{\theta\theta,T}(\theta_0) T^{1/2} L_{\theta,T}(\theta_0) + \zeta_T. \end{aligned} \tag{8}$$

It is easy to verify that $\zeta_T \stackrel{a.s.}{\to} 0$ as $T \to \infty$, as it can be written as

$$\begin{aligned} \zeta_T = &- \left(\tilde{L}^{-1}_{\theta\theta,T}(\bar{\theta}_T) - L^{-1}_{\theta\theta,T}(\bar{\theta}_T) \right) \left(T^{1/2} \tilde{L}_{\theta,T}(\theta_0) - T^{1/2} L_{\theta}(\theta_0) \right) \\ &- L^{-1}_{\theta\theta,T}(\bar{\theta}_T) \left(T^{1/2} \tilde{L}_{\theta,T}(\theta_0) - T^{1/2} L_{\theta,T}(\theta_0) \right) - T^{1/2} L_{\theta,T}(\theta_0) \left(\tilde{L}^{-1}_{\theta\theta,T}(\theta_0) - L^{-1}_{\theta\theta,T}(\theta_0) \right) \end{aligned}$$

where the first row converges a.s. to zero by Lemma C3. The second term in the second row converges by Lemma C3 and Proposition 1. The first term on the second row converges by Lemma C3 and, ensured by the continuity of $\ell(\theta)$ around θ_0, $L^{-1}_{\theta\theta,T}(\bar{\theta}_T)$ converges a.s. to $-\ell(\theta_0)^{-1}$.

Mean value expansion of the feasible criterion function around θ_0 gives

$$\tilde{g}_{i,t}(\tilde{\theta}_T) = \tilde{g}_{i,t}(\theta_0) + \tilde{g}_{i,\theta,t}(\bar{\theta}_T)(\tilde{\theta}_T - \theta_0).$$

Summing over t and scaling with $(T-m)^{-1/2}$ and using (8) gives

$$\begin{aligned}
\tilde{q}_{i,T}(\tilde{\theta}_T) &= T^{-1/2} \sum_{t=m+1}^{T} \tilde{g}_{i,t}(\theta_0) + (T-m)^{-1/2} \sum_{t=m+1}^{T} \tilde{g}_{i,\theta,t}(\bar{\theta}_T)(\tilde{\theta}_T - \theta_0) \\
&= T^{-1/2} \sum_{t=m+1}^{T} g_{i,t}(\theta_0) - \mathcal{G}_i(\theta_0) L^{-1}_{\theta\theta,T}(\theta_0) T^{1/2} L_{\theta,T}(\theta_0) + \zeta^\star_T \\
&= \begin{bmatrix} -\mathcal{G}_i(\theta_0) L^{-1}_{\theta\theta,T}(\theta_0) & I_m \end{bmatrix} T^{-1/2} \sum_{t=1}^{T} \xi_{i,t}(\theta_0) + \zeta^\star_T \xrightarrow{d} N(0, \Omega_i),
\end{aligned}$$

where $\zeta^\star_T \overset{a.s.}{\to} 0$ as $T \to \infty$, as Lemma D3 and the continuity of $g_{i,\theta,t}(\theta)$ in Θ_0 around θ_0 has been exploited (Lemma D1). The asymptotic covariance matrix is given by

$$\Omega_i = \begin{bmatrix} -\mathcal{G}_i(\theta_0)\mathcal{J}(\theta_0)^{-1} & I_{m\times m} \end{bmatrix} \Upsilon_i(\theta_0) \begin{bmatrix} -\mathcal{J}(\theta_0)^{-1}\mathcal{G}_i(\theta_0)' \\ I_{m\times m} \end{bmatrix}.$$

The χ^2_m-distribution follows using standard arguments, since the covariance matrices are of the rank m for all $m < T,\ m \in \mathbb{N}$. □

In Appendix E, it has been stated that the covariance matrix Ω_{hs} simplifies substantially, as it can be shown that the matrix $\mathcal{G}_{hs}(\theta_0) = 0_{m\times m}$. It follows that the heteroskedasticity test statistic is invariant to the estimation uncertainty. This result is in line with the findings in McLeod and Li (1983) in the conventional invertible case, and it makes the execution of the test easier in practice. The estimation uncertainty is not transmitted to the distribution of $Q_{hs,T}$. The distribution is solely determined by the properties of the true error process, not by the properties of its estimators.

Remark 1. *Given the limiting covariance matrix $\Omega_{hs,T}$ in Appendix E, the consistent estimator $\tilde{\Omega}_{hs,T}$ we suggest is*

$$\tilde{\Omega}_{hs,T} = \left(T^{-1} \sum_{t=1}^{T} \left(\tilde{u}_t(\tilde{\theta}_T)^2 - \tilde{\sigma}^2_T \right)^2 \right)^2 I_{m\times m}$$

It is easy to see that with this choice of $\tilde{\Omega}_{hs,T}$, the $Q_{hs,T}$ test statistic coincides numerically with the McLeod-Li portmanteau Q statistic (McLeod and Li, 1983), as it can be written as

$$Q_{hs,T} = (T-m) \sum_{j=1}^{m} \rho(j)^2, \quad \textit{with } \rho(j) = \frac{(T-m)^{-1} \sum_{t=m+1}^{T} \left(\tilde{u}_t(\tilde{\theta}_T)^2 - \tilde{\sigma}^2_T \right) \left(\tilde{u}_{t-j}(\tilde{\theta}_T)^2 - \tilde{\sigma}^2_T \right)}{T^{-1} \sum_{t=1}^{T} \left(\tilde{u}_t(\tilde{\theta}_T)^2 - \tilde{\sigma}^2_T \right)^2}.$$

4 Monte Carlo Simulations

4.1 Size simulations

In this section we study the finite sample properties of the proposed test statistics by Monte Carlo simulations. We begin with size simulations using two different data generating processes,

$$y_t = 0.2y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t+1} \quad \text{and} \tag{9}$$

$$y_t = 0.2y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t+1}. \tag{10}$$

Throughout this exercise we set $\sigma^2 = 2$ and $m = 5$. Error process η_t follows the Student's t-distribution with degrees of freedom λ_0. We vary the sample size as $T = 250$, 500, and 10,000. The smallest size $T = 250$ represents a magnitude often encountered when, for example, quarterly data is used. This is the case in our empirical example in the next section, where we use quarterly stock return data from 1947Q1 to 2007Q4. $T = 500$ is also relevant for lower frequency financial and macro economic data, for example when monthly data is under consideration for a shorter period of time. The largest sample size is used to illustrate the asymptotic properties of the statistics.

We use three different degrees of freedom, $\lambda_0 = 3$, 5, and 9, for the Student's t-distribution. Assumption 1 in Appendix A lays down the conditions on the moments of the innovations ε_t: Asymptotic properties of the $Q_{ac,T}$ test and the $Q_{hs,T}$ test have been derived under the assumption of finite fourth moments and finite eight moments, respectively. These assumptions are satisfied for $\lambda_0 > 4$ for the $Q_{ac,T}$ test, and for $\lambda_0 > 8$ in the case of the $Q_{hs,T}$ test. Our selection of the degrees of freedom parameters allows us to study the properties of the tests when the assumptions are met, but also illustrate how deviations from these conditions affect the properties of the tests. For $\lambda_0 = 3$, moment conditions fail to hold for both of the test statistics, and the deviation from this assumption is more severe for the $Q_{hs,T}$ test. If $\lambda_0 = 5$, the condition of $Q_{ac,T}$ test is satisfied, but that of the $Q_{hs,T}$ test is not, whereas $\lambda_0 = 9$ meets with the assumptions of both tests.

For each combination of the parameter values, we simulate 1,000 data sets using (9) and (10). To avoid initialization effects, 2,000 extra observations in the beginning of

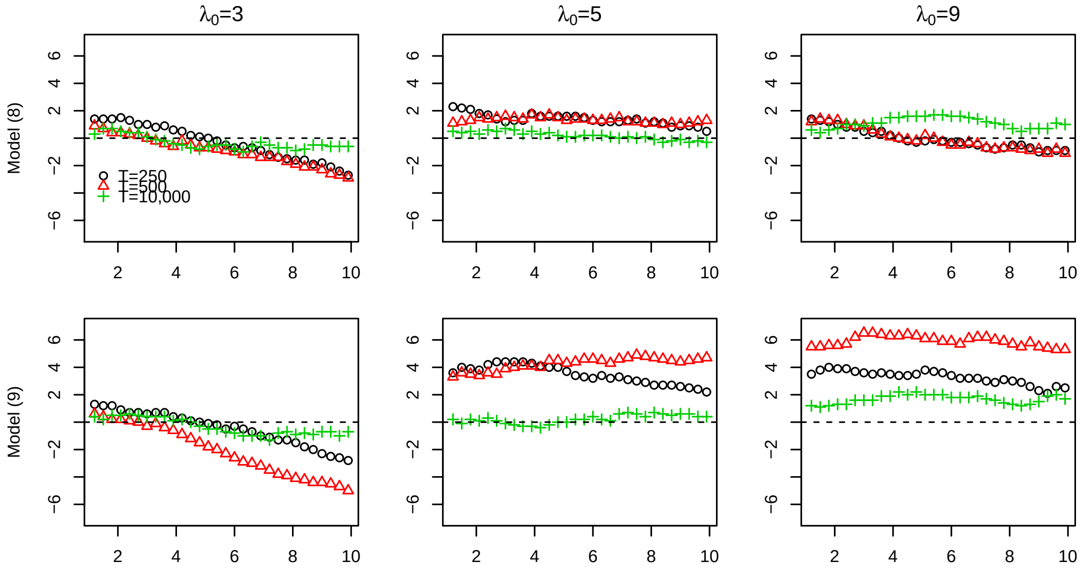

Figure 1: Size of the Q_{ac} test statistic for different sample sizes. The discrepancy (the size of the test minus the nominal size) is plotted against the nominal size of the test. Results are presented in percentage points. Different columns correspond to different values of λ_0 and rows to Model (9) and Model (10) respectively.

each series is simulated and discarded. To each of the series, we fit the noninvertible ARMA(1,1) model and use the residuals to perform the tests.

Summary of the simulation results is illustrated here graphically. [5] Figure 1 plots the discrepancy of the $Q_{ac,T}$ test size: the deviation of the tests actual size from its nominal size is plotted against the nominal size for significance levels $1\%, 1.2\%, \ldots 10\%$. Columns in the Figure 1 refers to different values of λ_0. A modest deviation from this assumption, $\lambda_0 = 3$, seems not very crucial, at least if the sample size is large enough. Test tends to overreject slightly for small significance levels and underreject for large significance levels. As the moment conditions are satisfied, the discrepancy is more evenly distributed across the significance levels, although for the modest sample sizes there is a tendency of slight overrejection (columns 2 and 3). Overrejection increases slightly as the MA parameter increases (second row). For the largest sample size, the $Q_{ac,T}$ test is, over all, rather well in line with its nominal size.

Figure 2 illustrates size discrepancies of the $Q_{hs,T}$ test statistic. The first column,

[5] Additional simulation results are available in the supplementary appendix, which is available upon request from the author. The size and power properties have been investigated using different parameter value combinations, and also Ljung-Box tests have been calculated for the sake of comparison.

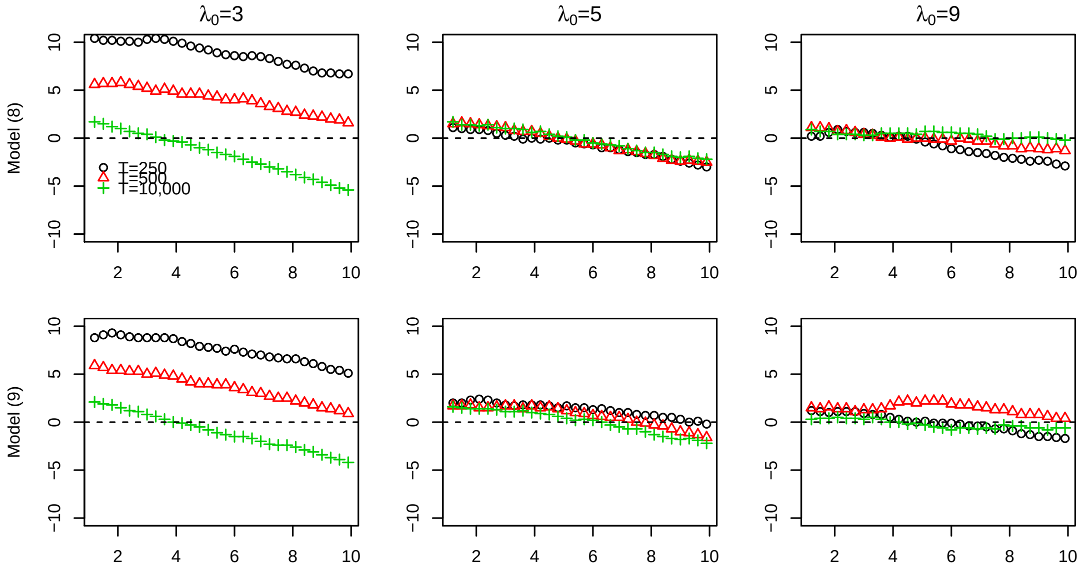

Figure 2: Size of the $Q_{hs,T}$ test statistic for different sample sizes. The discrepancy (the size of the test minus the nominal size) is plotted against the nominal size of the test. Results are presented in percentage points. Different columns correspond to different values of λ_0 and rows to Model (9) and Model (10) respectively.

where $\lambda_0 = 3$, points out the importance of the moment condition $\lambda_0 > 8$; the severe violation of the moment condition makes the χ^2-approximation of the Q_{hs} test statistics distribution much less accurate, even for the very large sample size. The approximation gets more accurate as the deviation from the moment conditions diminishes, as can be seen in the middle column where $\lambda_0 = 5$. For $\lambda_0 = 9$, the moment condition is fulfilled, and the size of the $Q_{hs,T}$ test is well approximated by the suggested χ^2-distribution.

4.2 Power simulations

Power properties of the tests are studied by simulating data using three different models that are more general than ARMA(1,1): one ARMA(1,2) model and two ARMA(1,1)-ARCH(1) models with different parameter values. The model equations are

$$y_t = 0.2y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t+1} - 0.2\varepsilon_{t+2}, \tag{11}$$

$$y_t = 0.2y_{t-1} + \sigma_t\eta_t - 0.2\sigma_{t+1}\eta_{t+1}, \quad \sigma_t = \sqrt{2 + 0.2\eta_{t-1}^2}, \quad \text{and} \tag{12}$$

$$y_t = 0.2y_{t-1} + \sigma_t\eta_t - 0.2\sigma_{t+1}\eta_{t+1}, \quad \sigma_t = \sqrt{2 + 0.8\eta_{t-1}^2}. \tag{13}$$

Model (11) is used to study the power of detecting misspecification of the lag length of the model (1). Models (12) and (13) are noninvertible ARMA(1,1) models with ARCH-

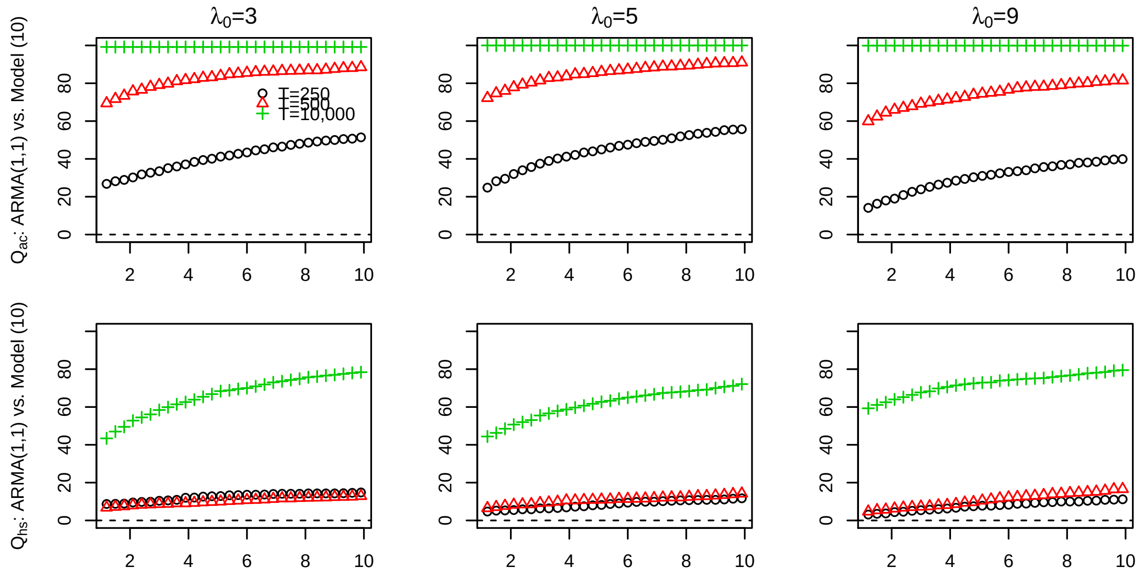

Figure 3: Power of the tests Q_{ac} (upper) and $Q_{hs,T}$ (lower) of ARMA(1,1) vs. Model (11) for different sample sizes. The size of the tests is plotted against the nominal size of the test, both measured in percentage points. Different columns correspond to different values of λ_0.

type heteroskedasticity, the processes studied in Meitz and Saikkonen (2013). The latter models are designed to illustrate the power of $Q_{hs,T}$ test against nonlinear models with heteroskedasticity.

The design of the Monte Carlo experiment is similar to the size simulations in the previous subsection. Again, we have simulated 1,000 data sets using models (10)-(12), and for each set, the noninvertible ARMA(1,1) model has been fitted and the test statistics have been calculated using the obtained feasible residuals.

Test statistics' power of detecting the misspecified lag length is illustrated in Figure 3. Autocorrelation in the residuals is well captured by the $Q_{ac,T}$ test, as can be see at the top row. Whenever the moment assumption $\lambda_0 > 4$ is satisfied (second and third column), the power lies between 75% and 85% for the 5% significance level, for the moderate sample size $T = 500$.

We would expect to find some heteroskedasticity in the residuals in this misspecification scenario, but it might be very mild. Therefore it is not an utter surprise that the $Q_{hs,T}$ test has limited power against it (bottom row in Figure 3). However, for very large sample sizes, the heteroskedasticity can be captured with moderate accuracy.

The $Q_{hs,T}$ tests capability of capturing a more severe type of heteroskedasticity, implied by the ARCH error term in Models (12) and (13), is illustrated in Figure 4. Test

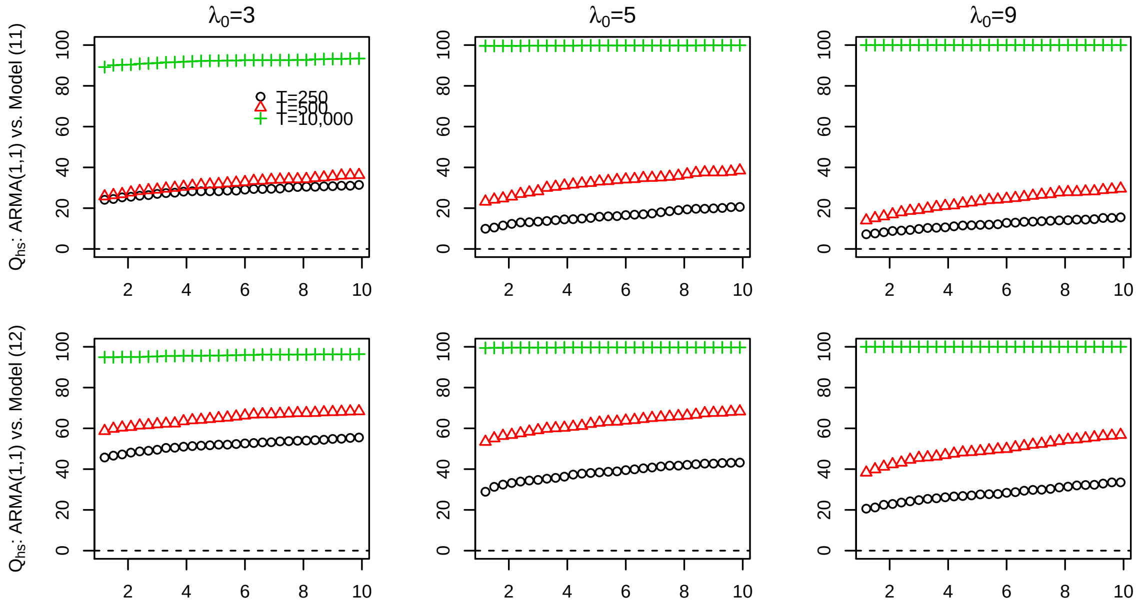

Figure 4: Power of the $Q_{hs,T}$ test of ARMA(1,1) vs. Model (11) (top) and Model (12) (bottom) for different sample sizes. The size of the tests is plotted against the nominal size of the test, both measured in percentage points. Different columns correspond to different values of λ_0.

performs well asymptotically, and the performance improves as the magnitude of the heteroskedasticity increases (bottom row).

5 Application to Financial Time Series Data

The question we address in this section is, if the diagnostic checks would shed light on the matter of predictability of asset returns. In our context, predictability simply means non-constant conditional expectation. According to dynamic asset pricing literature, predictability is a consequence of agents' risk aversion. For a thorough discussion, see Chapter 9 in Campbell, Lo, and MacKinlay (1997), or Chapter 2 in Singleton (2009). By no means should the predictability be manifested in a form of autocorrelation, but rather we are expecting to encounter nonlinear predictability.

The advantage the noninvertible model has over the invertible one in modeling asset returns is the generality of it. In the previous analysis of predictability, testing has usually been based on the invertible and autocorrelated ARMA model, which is implied, for example, by the price-trend model of Taylor (1982) or the mean-reversion model of Poterba and Summers (1988). The noninvertible ARMA model is capable of capturing all the same autocorrelaiton structures that the invertible model is, but also controlling for the nonlinearities that are often encountered in the financial time series data. As for the invertible ARMA model, the lack of autocorrelation automatically implies independence of the data, for the noninvertible ARMA model zero-autocorrelation is just a special case and the observations may still be dependent in a nonlinear fashion. This kind of generality allows us to model a richer class of dependencies with noninvertible model, than a conventional invertible ARMA model would allow.[6]

Following Lanne et al. (2013), we suggest that the noninvertible ARMA(1,1) model is particularly potential candidate in capturing this nonlinear predictability. To our knowledge, this is the first time the model is investigated from the standpoint of model fit based on asymptotic results. Preceding related work has mainly illustrated how the noninvertible model can mimic the nonlinear behavior of stock markets (Breidt et al., 2001), or how the predictability can be tested under the null of noninvertible ARMA model (Lanne et al., 2013). The evaluation of the model fit has been done so far merely by looking at the sample autocorrelation functions, without having the correct critical values.

Using statistical tests on the estimated parameters of the noninvertible ARMA(1,1) model, Lanne et al. (2013) reported nonlinear predictability, in line with the asset pricing theory. Their testing procedure implicitly assumed that under the null the correct model is the noninvertible ARMA. We take another look at this data and show that our diagnostic checks actually support this assumption and thus give support to their conclusions of nonlinear predictability.

In this section we apply our test statistics to evaluate the fit of the noninvertible ARMA models to the quarterly measured stock portfolio returns compiled of U.S. stocks. We use three value-weighted, size ordered stock portfolios, and the market portfolio, which include data from NYSE, AMEX, and NASDAQ stocks from January 1947 to December 2007, the same data that was used by Lanne et al. (2013). Data is obtained from

[6] An interesting special case of the noninvertible ARMA model (1) is one where roots of the AR polynomial coincide with the reciprocals of roots of the MA polynomial. It can be shown that for this so-called all-pass model the autocorrelation function will be zero for all lags, but the data is not iid. More generally, the squared observations of the noninvertible ARMA model (1) can be shown to be always autocorrelated, as long as there are nonzero AR and MA parameters. (See Appendix A.2 in Lanne et al. (2013).)

Kenneth French's web site.[7] Monthly returns are transformed into quarterly quantities by continuous compounding and means are subtracted from the series.

Estimation results are gathered in Table 1. The left side of the table shows the estimated parameters. It is worth noticing that the parameters are estimated with a good precision, they are statistically different from zero, and AR and MA parameters are close to each other. This suggests that the series are very mildly autocorrelated, but dependent some nonlinear way. Estimation has been based on the Student's t-distribution. The estimates of the degrees of freedom parameter λ_0 suggest that the innovation processes in all of the cases have finite fifth moments. This is enough to satisfy the moment condition imposed to the $Q_{ac,T}$ test, but it fails to meet the assumption of the finite eight moment of the $Q_{hs,T}$ test. Nevertheless, the Monte Carlo experiment in the previous section encourages us to still carry out the tests, with caution, as the size properties of the test were not too distorted by this relatively modest deviation from the moment condition.

The columns on the right give the p-values of the $Q_{ac,T}$ and $Q_{hs,T}$ tests for three different lag lengths, $m = 5$, 9, and 12. For three out of four portfolios, the null can not be rejected, suggesting that there is no autocorrelation left in the residuals or squared residuals. The noninvertible ARMA(1,1) model seems like an adequate model for the Market, Middle 40% and Top 30% portfolios in the light of our checks. The heteroskedasticity in the residuals of the noninvertible ARMA(1,1) model for Bottom 30% portfolio can not be ruled out, but it turns out that the noninvertible ARMA(2,2) model is suitable in controlling for that. (All the estimated parameters of the noninvertible ARMA(2,2) model are statistically highly significant (all p-values $< .01$) and the p-values of the $Q_{ac,T}$ and $Q_{hs,T}$ tests with $m = 5$ are 0.783 and 0.143 for the $Q_{ac,T}$ and $Q_{hs,T}$ tests, respectively, and similar for different choices of m as well.)

6 Conclusions

In this article we derived asymptotic properties for two residual-based test statistics for evaluating model adequacy of the noninvertible ARMA model. The $Q_{ac,T}$ test statistic is designed to detect remaining autocorrelation in the residuals and it is analogous to the

[7] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, downloaded Jan. 10, 2017

					$Q_{ac,T}$			$Q_{hs,T}$		
Portfolio	a	b	σ^2	λ	5	9	12	5	9	12
Market	.748 (.083)	.759 (.090)	8.074 (.673)	5.012 (1.803)	.982	.911	.961	.606	.845	.794
Bottom 30%	.846 (.039)	.936 (.037)	11.855 (.969)	5.285 (2.519)	1.000	.959	.942	.001	.001	.000
Middle 40%	.684 (.093)	.780 (.092)	9.826 (.751)	5.404 (2.117)	.880	.861	.924	.304	.376	.577
Top 30%	.746 (.081)	.721 (.092)	7.679 (.603)	5.152 (1.842)	1.000	.931	.963	.501	.850	.856

Table 1: The noninvertible ARMA(1,1) model has been estimated to four stock return index series. Table indicates the parameter estimates and their standard errors. Test statistics $Q_{ac,T}$ and $Q_{hs,T}$ have been calculated from the residuals of the fitted models and their p-values have been reported for three different lag lengths m for each test.

Box-Pierce Q-statistic. The asymptotic distribution of this test is not invariant to the estimation uncertainty of the model, so it must be taken into account in construction of the test. The $Q_{hs,T}$ test statistic is designed to capture autocorrelation in the squared residuals, and detect possible heteroskedasticity in the residuals. This test, in turn, is invariant to the estimation uncertainty, so we found that the McLeod-Li Q-statistic is asymptotically valid test for this purpose among the noninvertible models as well. Both tests have an asymptotic χ^2-distribution. Tests are simple to apply in practice, since the model needs to be estimated only under the null of correctly specified model and it has power against wide variety of misspecifications.

Our empirical example was designed to evaluate adequacy of the noninvertible ARMA model to the quarterly U.S. stock return data. The model was found, in light of our tests, a potential candidate in modeling these mildly autocorrelated, but possibly nonlinearly dependent data which, in turn, provides good grounds for looking for nonlinear predictability in the asset returns. Work in this direction has recently been done by Lanne et al. (2013), where the noninvertibility was implicitly assumed in their testing procedure. Our findings thus lend support to their assumption and moreover to their conclusions of possible nonlinear predictability.

In this article we based the asymptotic properties of the tests to the estimated model obtained by ML method. We do note, that there are other possible estimation methods available as well, for example the absolute deviation method by Breidt et al. (2001) and Wu and Davis (2010). These methods may lack some of the efficiency the ML method has, but there are certain benefits of not having to define the error distribution. We also note that the moment conditions for the tests, especially for the $Q_{hs,T}$ test, may prove to be heavy in practice. Although size and power were adequate even under moderate deviation from this assumption, there are other methods for testing hypothesis on residuals where the moment conditions may be aesed, such as the generalized spectral density by Hong (1999). These extensions are outside the scope of this article.

References

Alessi, L., M. Barigozzi, and M. Capasso (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review 79*, 16–47.

Andrews, B., M. Calder, and R. A. Davis (2009). Maximum likelihood estimation for α-stable autoregressive processes. *Annals of Statistics 37*, 1946–1982.

Andrews, B., R. A. Davis, and F. J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis 97*, 1638 – 1659.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 117*, 1329–1368.

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*, 1509–1526.

Breidt, F. J., R. A. Davis, and A. A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *Annals of Statistics 29*, 919–946.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*, Volume 2. Princeton University Press, Princeton, NJ.

Cui, Y., T. J. Fisher, and R. Wu (2014). Diagnostic tests for non-causal time series with infinite variance. *Journal of Statistical Planning and Inference 147*, 117–131.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007). ABCs (and Ds) of understanding VARs. *American Economic Review 97*, 1021–1026.

Forni, M. and L. Gambetti (2014). Sufficient information in structural VARs. *Journal of Monetary Economics 66*, 124–136.

Gouriéroux, C. and J.-M. Zakoïan (2013). Explosive bubble modelling by noncausal process. Working Papers 2013-04, Centre de Recherche en Economie et Statistique.

Hansen, L. P. and T. J. Sargent (1981). Exact linear rational expectations models: Specification and estimation. Technical report, Federal Reserve Bank of Minneapolis.

Hansen, L. P. and T. J. Sargent (1991). Two difficulties in interpreting vector autoregressions. In L. P. Hansen, T. J. Sargent, J. Heaton, A. Marcet, and W. Roberds (Eds.), *Rational expectations econometrics*. Oxford: Westview Press Boulder, CO.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association 94*, 1201–1220.

Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of non-invertible moving average processes. *Scandinavian Journal of Statistics 27*, 689–702.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014). Heterogeneous beliefs and tests of present value models. *Review of Economic Studies 81*, 1137–1163.

Lanne, M., M. Meitz, and P. Saikkonen (2013). Testing for linear and nonlinear predictability of stock returns. *Journal of Financial Econometrics 11*, 682–705.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*, 1115–1145.

Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for nongaussian nonminimum phase ARMA sequences. *Statistica Sinica 6*, 1–22.

Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*, 297–303.

Lof, M. (2013). Noncausality and asset pricing. *Studies in Nonlinear Dynamics & Econometrics 17*, 211–220.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis 114*, 227–255.

Nyberg, H., M. Lanne, and E. Saarinen (2012). Does noncausality help in forecasting economic time series? *Economics Bulletin 32*, 2849–2859.

Poterba, J. M. and L. H. Summers (1988). Mean reversion in stock prices. *Journal of Financial Economics 22*, 27 – 59.

Rosenblatt, M. (2012). *Gaussian and non-Gaussian linear time series and random fields.* Springer Science & Business Media.

Scott, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability 5*, 119–137.

Singleton, K. J. (2009). *Empirical dynamic asset pricing: model specification and econometric assessment*. Princeton University Press.

Taylor, S. J. (1982). Tests of the random walk hypothesis against a price-trend hypothesis. *Journal of Financial and Quantitative Analysis 17*, 37–61.

Wu, R. and R. A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis 31*, 98–112.

Appendix A Assumptions

This section lists the assumptions of the model that are used in the main text to derive the properties of the proposed test statistics. Let us develop some notation. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}_t$ a σ-algebra generated by the random variables $\{\eta_s\}_{s\leq t}$. The true but unknown parameters of the model are $\theta_{0,a} = (a_{0,1}, \dots, a_{0,P})$, $\theta_{0,b} = (b_{0,1}, \dots, b_{0,Q})$, σ_0 and $\lambda_0 = (\lambda_{0,1}, \dots, \lambda_{0,d})$. Let us collect these parameters into a $P+Q+1+d$ dimensional parameter vector $\theta_0 = (\theta_{0,a}, \theta_{0,b}, \sigma_0, \lambda_0)$. Furthermore, we define polynomials $a(z,\theta) = 1 - a_1 z - \cdots - a_P z^P$, and $b(z^{-1}, \theta) = 1 - b_1 z^{-1} - \cdots - b_Q z^{-Q}$ and gather the parameters in vectors $\theta_a = (a_1, \dots, a_P)$, $\theta_b = (b_1, \dots, b_Q)$ and $\theta = (\theta_a, \theta_b, \sigma, \lambda)$. Parameter vector θ defines the counterpart of the Model (1) for the parameter values $\theta \neq \theta_0$.

The partial derivative of the distribution function $f_\eta(x;\lambda)$ is denoted by sub-indices, $f_{\eta,x}(x;\lambda) = \frac{\partial}{\partial x} f_\eta(x;\lambda)$, $f_{\eta,\lambda}(x;\lambda) = \frac{\partial}{\partial \lambda} f_\eta(x;\lambda)$ and $f_{\eta,yz} = \frac{\partial^2}{\partial y \partial z} f_\eta(x;\lambda)$ where $y, z \in \{x, \lambda\}$. When the derivatives of the log-likelihood are considered, we use shorthand notations

$$e_{x,t}(\theta) = \frac{f_{\eta,x}\left(\sigma^{-1}u_t(\theta);\lambda\right)}{f_\eta\left(\sigma^{-1}u_t(\theta);\lambda\right)} \quad \text{and} \quad e_{\lambda,t}(\theta) = \frac{f_{\eta,\lambda}\left(\sigma^{-1}u_t(\theta);\lambda\right)}{f_\eta\left(\sigma^{-1}u_t(\theta);\lambda\right)},$$

where the subscript denotes the partial derivative of the density function, $f_{\eta,x}(x;\lambda) = \frac{\partial}{\partial x} f_\eta(x;\lambda)$.

The first assumption summarizes the restrictions imposed on the error process ε_t.

Assumption 1. *The error process is $\varepsilon_t = \sigma_0 \eta_t$ with η_t an iid sequence with $E[\eta_t] = 0$ and $E[\eta^2] = 1$. The distribution of η_t is symmetric and non-Gaussian with density function $f_\eta(x;\lambda_0)$, where λ_0 is a $d \times 1$ parameter vector. In addition to the finite second moments, the process η_t has either*

(a) finite fourth moments, $E[\eta_t^4] < \infty$, or

(b) finite eight moments, $E[\eta_t^8] < \infty$.

The permissible parameter space Θ is one that satisfies the causality and invertibility conditions for the polynomials $a(z,\theta)$ and $b(z^{-1},\theta)$, and the assumptions of the positive and finite error term variance:

Assumption 2. *The permissible parameter space is* $\Theta = \Theta_a \times \Theta_b \times \Theta_\sigma \times \Theta_\lambda$*, where*

$$\begin{aligned} \Theta_a &= \left\{\theta_a \in \mathbb{R}^P \ ; \ a(z) \neq 0 \ \forall \ |z| \leq 1\right\}, \\ \Theta_b &= \left\{\theta_b \in \mathbb{R}^Q \ ; \ b(z^{-1}) \neq 0 \ \forall \ |z| \leq 1\right\}, \\ \Theta_\sigma &= \left\{\sigma \in \mathbb{R}^+\right\} \qquad \textit{and} \\ \Theta_\lambda &= \{\lambda \in \mathbb{R}^d\}. \end{aligned}$$

The true parameter θ_0 *lies in a compact and convex set* $\Theta_0 \subset \Theta$*.*

The following high level assumptions concerns the distribution of the innovation process η_t, and can be found in Meitz and Saikkonen (2013).

Assumption 3.

A1. (*i*) *For all* $x \in \mathbb{R}$ *and* $\lambda \in \Theta_\lambda$*,* $f_\eta(x;\lambda)$ *is twice continuously differentiable w.r.t.* (x,λ)*.*

(*ii*) *For all* $\lambda \in \Theta_\lambda$*,* $\int x f_\eta(x;\lambda)dx = 0$ *and* $\int x^2 f_\eta(x;\lambda)dx = 1$*.*

(*iii*) *The matrix* $E[e_{\lambda,t}(\theta_0)e_{\lambda,t}(\theta_0)']$ *is positive definite.*

(*iv*) *For all* $x \in \mathbb{R}$ *and all* λ_i*,* $i = 1, \dots, d$*, the functions*

$$x^4 \frac{f^2_{\eta,x}(x;\lambda_0)}{f^2_\eta(x;\lambda_0)} \quad \textit{and} \quad \frac{f^2_{\eta,\lambda_i}(x;\lambda_0)}{f^2_\eta(x;\lambda_0)}$$

are dominated by $d_1(1+|x|^{d_2})$ *with some* $d_1, d_2 \geq 0$ *s.t.* $\int |x|^{d_2} f_\eta(x;\lambda_0)dx < \infty$*.*

(*v*) *For all* $x \in \mathbb{R}$ *and* $\lambda \in \Theta_\lambda$*, the function* $|x^2 f_{\eta,\lambda}(x;\lambda)|$ *is dominated by function* $\bar{f}(x)$ *s.t.* $\int \bar{f}(x)dx < \infty$*.*

A2. (*i*) *For all* $x \in \mathbb{R}$ *and* $\lambda \in \Theta_\lambda$*, the function* $|f_{\eta,\lambda\lambda}(x;\lambda)|$ *is dominated by some* $\bar{f}(x)$ *s.t.* $\int \bar{f}(x)dx < \infty$*.*

(*ii*) $\int f_{\eta,xx}(x;\lambda_0)dx = 0$*.*

(*iii*) $\int x^2 f_{\eta,xx}(x;\lambda_0)dx = 2$*.*

A3. (*i*) *For all* $x \in \mathbb{R}$ *and* $\lambda \in \Theta_\lambda$*, for all* λ_i*,* $i = 1, \dots, d$*, the functions*

$$x^4 \frac{f^4_{\eta,x}(x;\lambda)}{f^4_\eta(x;\lambda)}, \quad \frac{f^4_{\eta,\lambda_i}(x;\lambda)}{f^4_\eta(x;\lambda)}, \quad x^4 \frac{f^2_{\eta,xx}(x;\lambda)}{f^4_\eta(x;\lambda)}, \quad \frac{f^2_{\eta,\lambda_i x}(x;\lambda)}{f^4_\eta(x;\lambda)}, \quad \textit{and} \quad \left|\frac{f_{\eta,\lambda\lambda}(x;\lambda)}{f_\eta(x;\lambda)}\right|$$

are dominated by $d_1(1+|x|^{d_2})$ *for some* $d_1, d_2 \geq 0$*, and* $\int |x|^{d_2} f_\eta(x;\lambda_0)dx < \infty$*.*

A4. (*i*) *For all $x \in \mathbb{R}$, $\Delta x \in \mathbb{R}$, and $\lambda \in \Theta_\lambda$, for some $C < \infty$ and $d_1, d_2 \geq 0$,*

$$|v(x + \Delta x; \lambda) - v(x; \lambda)| \leq C\left((1 + |x|^{d_1})|\Delta x| + |\Delta x|^{d_2}\right)$$

for the following functions $v(x; \lambda)$,

$$\begin{aligned}
&(i)\ v(x;\lambda) = \frac{f_{\eta,x}(x;\lambda)}{f_\eta(x;\lambda)}, \quad (ii)\ v(x;\lambda) = \frac{f_{\eta,\lambda}(x;\lambda)}{f_\eta(x;\lambda)},\\
&(iii)\ v(x;\lambda) = \frac{f_{\eta,\lambda\lambda}(x;\lambda)}{f_\eta(x;\lambda)}, \quad (iv)\ v(x;\lambda) = \frac{f_{\eta,\lambda x}(x;\lambda)}{f_\eta(x;\lambda)} \quad \text{and}\\
&(v)\ v(x;\lambda) = \frac{f_{\eta,\lambda\lambda}(x;\lambda)}{f_\eta(x;\lambda)}.
\end{aligned}$$

Assumptions 1, 2, and 3 here are enough to state the following properties of (y_t, ε_t).

Lemma A1. *Under Assumptions 1 (a) and 2, the process (y_t, ε_t) defined in* (1) *is stationary and ergodic, and moreover, process y_t is $\mathcal{F}_{t+Q}$-measurable with $E[y_t^4] < \infty$, and ε_t is $\mathcal{F}_t$-measurable with $E[\varepsilon_t^4] < \infty$. In addition, under Assumption 1 (b), $E[y_t^8] < \infty$ and $E[\varepsilon_t^8] < \infty$.*

The proof will be omitted here, but essentially it can be found in Meitz and Saikkonen (2013) Appendix A, where the series presentations of rational functions like $a(z,\theta)^{-1}$, $b(z^{-1},\theta)^{-1}$, $a(z,\theta)^{-1}b(z^{-1},\theta)$ and $a(z,\theta)b(z^{-1},\theta)^{-1}$ are discussed in depth. For future reference, we list the definitions of these sums here:

$$\begin{aligned}
&a(z,\theta)^{-1} = \sum_{j=0}^{\infty} \psi_j^{(a)} z^j, \quad b(z^{-1},\theta)^{-1} = \sum_{j=0}^{\infty} \psi_j^{(b)} z^{-j}\\
&a(z,\theta)^{-1}b(z^{-1},\theta) = \sum_{j=-P}^{\infty} \psi_j z^j \quad \text{and} \quad a(z,\theta)b(z^{-1},\theta)^{-1} = \sum_{j=-Q}^{\infty} \pi_j z^{-j}.
\end{aligned}$$

These series expansions are well defined for all z in some area containing the unit circle, and the coefficients of the expansions are always geometrically decaying for all $\theta \in \Theta$.

Appendix B Derivatives of $u_t(\theta)$ and $\tilde{u}_t(\theta)$

The sequence $u_t(\theta)$ was defined in (3) and its feasible counterpart $\tilde{u}_t(\theta)$ in (4). For what follows, we need a notion of the derivatives of these quantities. The derivatives of $u_t(\theta)$ w.r.t. the p^{th} AR parameter and q^{th} MA parameter are denoted by $u_{a_p,t}(\theta) = \frac{\partial}{\partial a_p} u_t(\theta)$

and $u_{b_q,t}(\theta) = \frac{\partial}{\partial b_q} u_t(\theta)$, respectively, for $p = 1, \ldots, P$ and $q = 1, \ldots, Q$. These functions are given by

$$u_{a_p,t}(\theta) = -\frac{u_{t-p}(\theta)}{a(B)} = -\sum_{j=0}^{\infty} \psi_j^{(a)} u_{t-p-j}(\theta) \quad \text{and}$$
$$u_{b_q,t}(\theta) = \frac{u_{t+q}(\theta)}{b(B^{-1})} = \sum_{j=0}^{\infty} \psi_j^{(b)} u_{t+q+j}(\theta).$$

We will also use the derivative functions of $\tilde{u}_t(\theta)$ in (4). Using representations of these functions by Andrews, Davis, and Breidt (2006),

$$\tilde{u}_t(\theta) = \sum_{j=0}^{T-t} \psi_j^{(b)} a(B) y_{t+j},$$

the derivatives can be written as[8]

$$\tilde{u}_{a_p,t}(\theta) = -\sum_{j=0}^{T-t} \psi_j^{(b)} y_{t-p+j} \quad \text{and} \quad \tilde{u}_{b_q,t}(\theta) = \sum_{j=0}^{T-t} \psi_j^{(b)} \tilde{u}_{t+q+j}(\theta).$$

For convenience, the $P \times 1$ and $Q \times 1$ derivative vectors of $u_t(\theta)$, w.r.t. the parameter vectors θ_a and θ_b are denoted by

$$\frac{\partial}{\partial \theta_a} u_t(\theta) = u_{a,t}(\theta) = \begin{bmatrix} u_{a_1,t}(\theta) \\ \vdots \\ u_{a_P,t}(\theta) \end{bmatrix} \quad \text{and} \quad \frac{\partial}{\partial \theta_b} u_t(\theta) = u_{b,t}(\theta) = \begin{bmatrix} u_{b_1,t}(\theta) \\ \vdots \\ u_{b_Q,t}(\theta) \end{bmatrix}$$

and respectively for the feasible quantities $\tilde{u}_t(\theta)$ in an obvious manner.

Appendix C Intermediate Results

In this section we provide some preliminary results that are used extensively in what follows. First, we set some further notation. Approximation of the log-likelihood function was already given in (5). Its feasible counterpart is

$$\tilde{L}_T(\theta) = T^{-1} \sum_{t=1}^{T} \tilde{l}_t(\theta) \quad \text{with} \quad \tilde{l}_t(\theta) = \log f_\eta \left(\frac{\tilde{u}_t(\theta)}{\sigma}; \lambda \right) - \log \sigma.$$

[8] For more details, see Meitz and Saikkonen (2013), Appendix E.

The $P+Q+1+d$ dimensional score vector of a single observation at θ is denoted by $l_{\theta,t}(\theta) = \frac{\partial}{\partial\theta} l_t(\theta)$, and it is

$$l_{\theta,t}(\theta) = \begin{bmatrix} e_{x,t}(\theta)\sigma^{-1}u_{a,t}(\theta) \\ e_{x,t}(\theta)\sigma^{-1}u_{b,t}(\theta) \\ -\dfrac{1}{\sigma}\left(e_{x,t}(\theta)\dfrac{u_t(\theta)}{\sigma}+1\right) \\ e_{\lambda,t}(\theta) \end{bmatrix}.$$

The score vector of the model is given by $L_{\theta,T}(\theta) = T^{-1}\sum_{t=1}^{T} l_{\theta,t}(\theta)$.

The Hessian of the noninvertible ARMA model is more involved than that of the invertible ARMA model (although it is simplified substantially from the Hessian presented in Meitz and Saikkonen (2013) as we neglect the ARCH error term). It is not shown here, but one can confirm that

$$-\mathrm{E}\left[l_{\theta\theta',t}(\theta_0)\right] = \lim_{T\to\infty} \mathrm{Cov}\left(T^{-1/2}\sum_{t=1}^{T} l_{\theta,t}(\theta_0)\right),$$

where the limit is a positive definite, continuous and finite in some neighborhood Θ_0 of θ_0. This matrix is (see Proposition 1)

$$\ell(\theta_0) = \begin{bmatrix} A_{11} & B_{21}' & 0_{P\times 1} & 0_{P\times d} \\ B_{21} & A_{22} & 0_{Q\times 1} & 0_{Q\times d} \\ 0_{1\times P} & 0_{1\times Q} & A_{33} & A_{43}' \\ 0_{d\times P} & 0_{d\times Q} & A_{43} & A_{44} \end{bmatrix}.$$

Straightforward but rather long calculations give the following expressions for the blocks:

$$\begin{aligned} A_{11} &= -\sigma_0^{-2}\mathrm{E}\left[e_{x,t}^2(\theta_0)\right]\mathrm{E}\left[u_{\theta_a,t}(\theta_0)u_{\theta_a',t}(\theta_0)\right], \\ A_{22} &= \sigma_0^{-2}\mathrm{E}\left[e_{x,t}^2(\theta_0)u_{\theta_b,t}(\theta_0)u_{\theta_b',t}(\theta_0)\right], \\ A_{33} &= \sigma_0^{-2}\left(\mathrm{E}\left[e_{x,t}^2(\theta_0)\eta_t^2\right]-1\right), \\ A_{43} &= -\sigma_0^{-1}\mathrm{E}\left[e_{x,t}(\theta_0)e_{\lambda,t}(\theta_0)\eta_t\right], \\ A_{44} &= -\mathrm{E}\left[e_{\lambda,t}(\theta_0)e_{\lambda',t}(\theta_0)\right], \end{aligned}$$

and the block B_{21} has the (p,q) element

$$-\sum_{j=0}^{\infty}\psi_{0,j-p}^{(a)}\psi_{0,j-q}^{(b)}.$$

The block denoted by B is due to the serial correlation of the score vector, whereas the blocks denoted by A captures the contemporaneous correlation. The expressions given above have feasible counterparts which are obtained by using

$$\tilde{e}_{x,t}(\theta_0) = \frac{f_{\eta,x}(\sigma^{-1}\tilde{u}_t(\theta_0);\lambda_0)}{f_\eta(\sigma_0^{-1}\tilde{u}_t(\theta_0);\lambda_0)}, \quad \tilde{e}^2_{\lambda,t}(\theta_0) = \frac{f_{\eta,\lambda}(\sigma^{-1}\tilde{u}_t(\theta_0);\lambda_0)}{f_\eta(\sigma_0^{-1}\tilde{u}_t(\theta_0);\lambda_0)}, \quad \tilde{u}_t(\theta_0) = \sum_{j=0}^{T-t}\psi_{0,j}^{(b)}a_0(B)y_{t+j},$$

$$\tilde{u}_{a_p,t}(\theta_0) = -\sum_{j=0}^{T-t}\psi_{0,j}^{(b)}y_{t-p+j}, \quad \text{and} \quad \tilde{u}_{b_q,t}(\theta_0) = \sum_{j=0}^{T-t}\psi_{0,j}^{(b)}\tilde{u}_{t+q+j}(\theta_0)$$

instead of their unfeasible counterparts.

The set of results in Lemma C1 is used by Meitz and Saikkonen (2013) to derive the result we presented in Proposition 1, but it is also needed in the proof of Lemma C2 and Lemmas D1-D3 and Lemmas E4 and E5 in the forthcoming sections. Moreover, Lemmas C1 and C2 are needed for the proof of Lemma C3, which in turn, is used directly in the proof of Theorem 1.

Lemma C1. *Under Assumptions 1 (a), 2, and 3,*

$$(i)\ \left\|\sup_{\theta\in\Theta_0}|u_t(\theta)|\right\|_4 < \infty, \quad (ii)\ \left\|\sup_{\theta\in\Theta_0}\left|\frac{u_t(\theta)}{a(B)}\right|\right\|_4 < \infty, \quad (iii)\ \left\|\sup_{\theta\in\Theta_0}\left|\frac{u_t(\theta)}{b(B^{-1})}\right|\right\|_4 < \infty$$

$$(iv)\ \left\|\sup_{\theta\in\Theta_0}|e_{\lambda_l,t}(\theta)|\right\|_2 < \infty, \quad (v)\ \left\|\sup_{\theta\in\Theta_0}|e_{x,t}(\theta)|\right\|_2 < \infty, \quad (vi)\ E[e_{x,t}(\theta_0)] = 0,$$

$$(vii)\ E[e_{x,t}(\theta_0)\varepsilon_t] = -\sigma_0, \quad (viii)\ E\left[e_{x,t}(\theta_0)\varepsilon_t^2\right] = 0, \quad (ix)\ E\left[e_{x,t}(\theta_0)\varepsilon_t^3\right] = -3\sigma_0^3,$$

$$(x)\ E[e_{\lambda,t}(\theta_0)] = 0, \quad (xi)\ E\left[\varepsilon_t^2 e_{\lambda,t}(\theta_0)\right] = 0.$$

Proof. Proofs for the results in this section are given in the supplementary appendix. □

The next Lemma provides some insight into the relation between the unfeasible and feasible quantities, and these results are used extensively to prove Lemma C3 below.

Lemma C2. *Under Assumptions 1 (a), 2 and 3,*

$$(i)\ \left\|\sup_{\theta\in\Theta_0}|u_t(\theta)-\tilde{u}_t(\theta)|\right\|_4 \le CK_t, \quad (ii)\ \left\|\sup_{\theta\in\Theta_0}\left|u_{a_p,t}(\theta)-\tilde{u}_{a_p,t}(\theta)\right|\right\|_4 \le CK_t,$$

$$(iii)\ \left\|\sup_{\theta\in\Theta_0}\left|u_{b_p,t}(\theta)-\tilde{u}_{b_p,t}(\theta)\right|\right\|_4 \le CK_t, \quad (iv)\ \left\|\sup_{\theta\in\Theta_0}|e_{x,t}(\theta)-\tilde{e}_{x,t}(\theta)|\right\|_{r_1} \le CK_t,$$

$$(v)\ \left\|\sup_{\theta\in\Theta_0}\left|u_t(\theta)^2-\tilde{u}_t(\theta)^2\right|\right\|_2 \le CK_t,$$

where $C<\infty$ is a constant and may vary from part to part, and K_t is a constant that may depend on t and decays to zero at geometric rate as $T\to\infty$, and r_1 is some small positive number.

The following results are used directly in the proof of Theorem 1.

Lemma C3. *Let η_t be iid with $E[\eta_t] = 0$ and $E[\eta_t^2] = 1$, then under Assumptions 2 and 3, as $T \to \infty$,*

$$(i) \quad T^{1/2} \sup_{\theta \in \Theta_0} \left| L_{\theta,T}(\theta) - \tilde{L}_{\theta,T}(\theta) \right| \overset{a.s.}{\to} 0,$$

$$(ii) \quad \sup_{\theta \in \Theta_0} \left| L_{\theta\theta,T}(\theta) - \tilde{L}_{\theta\theta,T}(\theta) \right| \overset{a.s.}{\to} 0.$$

Appendix D Intermediate Properties for the Test Statistics

The next three lemmas are used repeatedly in the proof of Theorem 1. The first one, Lemma D1, gives some essential uniform convergence results for the quantities in the test statistics. Lemma D2 elaborates on the asymptotic distribution of the test statistics. A set of sufficient conditions for the asymptotic normality of $(T-m)^{-1/2} \sum_{t=m+1}^{T} g_{i,t}(\theta_0)$ is given, and the conditions are shown to hold. Lemma D3 considers the issue of the feasibility of the quantities. When it comes down to the asymptotics of the feasible test statistics, Lemma D3 with Lemma C3 gives the results that ensure that it does not matter if the feasible quantities are used instead of the unobservable unfeasible ones when deriving the asymptotic properties of the test statistics.

Lemma D1. *Under Assumptions 1 (a), 2 and 3, and for $i = ac$ and $g_{i,t}(\theta)$ as in (7),*

$$(i) \quad E[g_{i,t}(\theta_0)] = 0,$$

$$(ii) \quad \sup_{\theta \in \Theta_0} \left| \frac{1}{T} \sum_{t=1}^{T} \tfrac{\partial}{\partial \theta'} g_{i,t}(\theta) - \mathcal{G}_i(\theta) \right| \to 0 \ a.s. \ as \ T \to \infty,$$

$$(iii) \quad \sup_{\theta \in \Theta_0} \left| \frac{1}{T} \sum_{t=1}^{T} g_{i,t}(\theta) g_{i,t}(\theta)' - \mathcal{H}_i(\theta) \right| \to 0 \ a.s. \ as \ T \to \infty$$

$$(iv) \quad \sup_{\theta \in \Theta_0} \left| \frac{1}{T} \sum_{t=1}^{T} g_{i,t}(\theta) l_{\theta,t}(\theta)' - \Psi_i(\theta) \right| \to 0 \ a.s. \ as \ T \to \infty$$

where $\mathcal{G}_i(\theta) = E\left[\tfrac{\partial}{\partial \theta'} g_{i,t}(\theta)\right] = E[g_{i,\theta,t}(\theta)]$, $\mathcal{H}_i(\theta) = E[g_{i,t}(\theta) g_{i,t}(\theta)']$ and $\Psi_i(\theta) = E[g_{i,t}(\theta) l_{\theta,t}(\theta)']$ are constant matrices that are finite and $\mathcal{H}_i(\theta)$ is positive definite. Under Assumption 1 (b) result holds also for $i = hs$.

Proof. The proofs for the results in this section are available in the supplementary appendix. □

In order to show the asymptotic normality for $\tilde{q}_{i,T}(\tilde{\theta}_T)$, we proceed, by applying the central limit theorem of L_2-mixingales (Scott, 1973) to the sequence $T^{-1/2}\sum_{t=1}^{T}\xi_{i,t}(\theta_0)$, where, for $i \in \{ac, hs\}$,

$$\xi_{i,t}(\theta_0) \overset{def}{=} \begin{bmatrix} l_{\theta,t}(\theta_0) \\ g_{i,t}(\theta_0) \end{bmatrix}.$$

This strategy is appealing, since we need to show the mixingale property only for the $\mathcal{F}_t$-measurable subvector $g_{i,t}(\theta_0)$, the result for the score vector $l_{\theta,t}(\theta_0)$ being already shown in Meitz and Saikkonen (2013). The mixingale property of the $\mathcal{F}_t$-measurable vector is trivial. Asymptotic normality of the score of noncausal or noninvertible ARMA model is more involved than that of the conventional causal and invertible ARMA model. This is due to the fact that the score, namely $u_{b_q,t}(\theta_0)$, is not $\mathcal{F}_t$ measurable, and it is an autocorrelated sequence.

Lemma D2 below provides a set of sufficient conditions under which the vector $T^{-1/2}\sum_{t=1}^{T}\xi_{i,t}(\theta_0)$ has an asymptotic normal distribution with a positive definite covariance matrix $\Upsilon_i(\theta_0)$. This distribution characterizes the uncertainty in the estimation of sample autocorrelation functions in our test statistics.

Lemma D2. *For $i \in \{ac, hs\}$, Let $g_{i,t}(\theta)$ be as in* (7)*, and let $\xi_{i,t}(\theta) = (l_{\theta,t}(\theta), g_{i,t}(\theta))$. Then, under Assumptions 1 (a), 2, and 3,*

(*i*) *vectors $\xi_{i,t}(\theta_0)$ forms a stationary and ergodic process with $E[\xi_{i,t}(\theta_0)] = 0$,*

(*ii*) *this vector has a finite covariance matrix, $E\left[\xi_{i,t}(\theta_0)\xi'_{i,t}(\theta_0)\right] < \infty$,*

(*iii*) *for all conformable nonrandom vectors $a \neq 0$, the sequence $a'\xi_{i,t}(\theta_0)$ is an L_2-mixingale of size -1 w.r.t. filtration $\{\mathcal{F}_s\}_{s\leq t}$, and*

(*iv*) *there is a finite and positive definite limiting covariance matrix $\Upsilon_i(\theta_0)$ s.t.*

$$Cov\left(T^{-1/2}\sum_{t=1}^{T}\xi_{i,t}(\theta_0)\right) \to \Upsilon_i(\theta_0) \quad a.s.\ as\ T \to \infty.$$

Proposition 2. *For $i \in \{ac, hs\}$, Let $g_{i,t}(\theta)$ be as in* (7), *and let $\xi_{i,t}(\theta) = (l_{\theta,t}(\theta), g_{i,t}(\theta))$. Then, under Assumptions 1* (a), *2, and 3,*

$$T^{-1/2} \sum_{t=1}^{T} \xi_{i,t}(\theta_0) \xrightarrow{d} N(0, \Upsilon_i(\theta_0)),$$

where the positive definite covariance matrix is given in Lemma D2 (iv).

Proof. An application of Lemma A.4 in Meitz and Saikkonen (2013) and Lemma D2 (i)-(iv) yields the result $T^{-1/2} \sum_{t=1}^{T} a'\xi_{i,t}(\theta_0) \xrightarrow{d} N(0, a'\Upsilon_i(\theta_0)a)$ for all conformable non-random vectors $a \neq 0$. Proposition 2 follows by the Cramér-Wold device.

□

The actual form of the asymptotic covariance matrix $\Upsilon_i(\theta_0)$ must be known in order to execute the diagnostic tests on residuals. It is also clear that the form depends on the test we are executing, namely the form of function $g_{i,t}(\cdot)$. Stationarity of the process $\xi_{i,t}(\theta_0)$ allows us to write it in a form that eases the calculation of the actual matrix a little bit:

$$\begin{aligned}
\operatorname{Cov}\left(T^{-1/2} \sum_{t=1}^{T} \xi_{i,t}(\theta_0)\right) =& \mathrm{E}\left[\xi_{i,t}(\theta_0)\xi'_{i,t}(\theta_0)\right] \\
&+ \sum_{j=1}^{T-1} \frac{T-j}{T} \mathrm{E}\left[\xi_{i,t}(\theta_0)\xi'_{i,t-j}(\theta_0) + \xi_{i,t}(\theta_0)\xi'_{i,t+j}(\theta_0)\right] \\
\xrightarrow{T\to\infty}& \sum_{s=-\infty}^{\infty} \mathrm{E}\left[\xi_{i,t}(\theta_0)\xi'_{i,t-s}(\theta_0)\right].
\end{aligned}$$

For now on, let us divide the covariance matrix into four blocks as

$$\Upsilon_i(\theta_0) = \begin{bmatrix} \ell(\theta_0) & \Upsilon'_{i,\Psi}(\theta_0) \\ \Upsilon_{i,\Psi}(\theta_0) & \Upsilon_{i,\mathcal{H}}(\theta_0) \end{bmatrix}, \tag{21}$$

where the upper left corner is now already familiar asymptotic covariance of the score (see Proposition 1). The off-diagonal blocks reflects the co-movements of our parameter estimates and the values of the sample autocorrelation coefficients. The bottom right block gives the covariance matrix for the sample autocorrelation functions of the true, but not observable error terms of the model.

Under the assumptions leading to the results in Lemmas D1 and D2 the asymptotic distributional results can be shown for the unfeasible quantities. The next lemma ensures that the results hold for the feasible quantities as well.

Lemma D3. *For $i \in \{ac, hs\}$, let $g_{i,t}(\theta)$ be as in (6), $g_{i,\theta,t}(\theta)$ its derivative w.r.t. θ, and let $\tilde{g}_{i,t}(\theta)$ and $\tilde{g}_{i,\theta,t}(\theta)$ denote their feasible counterparts. Under Assumptions 1 (a), 2, and 3, we have the following uniform convergences,*

$$(i)\ \sup_{\theta\in\Theta_0}\left|T^{-1/2}\sum_{t=1}^{T} g_{i,t}(\theta) - T^{-1/2}\sum_{t=1}^{T}\tilde{g}_{i,t}(\theta)\right| \longrightarrow 0 \text{ a.s. as } T\to\infty \quad \text{and}$$

$$(ii)\ \sup_{\theta\in\Theta_0}\left|T^{-1}\sum_{t=1}^{T} g_{i,\theta,t}(\theta) - T^{-1}\sum_{t=1}^{T}\tilde{g}_{i,\theta,t}(\theta)\right| \longrightarrow 0 \text{ a.s. as } T\to\infty.$$

This completes the list of intermediate results for the proof of Theorem 1.

Appendix E Covariance Matrices Ω_{ac} and Ω_{hs}

In this Section, we give the exact forms of the limiting covariance matrices Ω_{ac} and Ω_{hs} in Theorem 1. Thorough derivations can be found in supplementary appendix. The matrix $\mathcal{J}(\theta_0)$ is defined in Proposition 1 and its form is given in Section C above.

In Theorem 1 we concluded that the limiting covariance matrices Ω_i, $i \in \{ac, hs\}$, are of the form

$$\Omega_i = \begin{bmatrix} -\mathcal{G}_i(\theta_0)\mathcal{J}(\theta_0)^{-1} & I_{m\times m}\end{bmatrix} \Upsilon_i(\theta_0) \begin{bmatrix} -\mathcal{J}(\theta_0)^{-1}\mathcal{G}_i(\theta_0)' \\ I_{m\times m}\end{bmatrix},$$

where the blocks are defined in Theorem 1 and discussion therein.

In the supplementary appendix, we have derived the exact forms of matrices $\mathcal{G}_i(\theta)$, $\Upsilon_{i,\mathcal{H}}(\theta)$ and $\Upsilon_{i,\Psi}(\theta)$ for all $\theta \in \Theta$. The exact form of Ω_i follows from these derivations by evaluating the matrices at θ_0 and using Lemma C1 $(vii) - (xi)$. Details are omitted here. After the exact forms are given, we discuss briefly the consistent estimation of these matrices.

Asymptotic covariance matrix Ω_{ac}

Matrix $\mathcal{G}_{ac}(\theta_0)$ can be written as a 1×4 block matrix

$$\mathcal{G}_{ac}(\theta_0) = \begin{bmatrix} \underset{m\times P}{\mathcal{G}_{ac,A}(\theta_0)} & \underset{m\times Q}{\mathcal{G}_{ac,B}(\theta_0)} & \underset{m\times 1}{\mathcal{G}_{ac,C}(\theta_0)} & \underset{m\times d}{\mathcal{G}_{ac,D}(\theta_0)}\end{bmatrix},$$

where the blocks have typical (k,l)-elements

$$[\mathcal{G}_{ac,A}(\theta_0)]_{k,l} = -\sigma_0^2\psi_{0,k-l}^{(a)} \quad \text{for } k \geq l,$$

$$[\mathcal{G}_{ac,B}(\theta_0)]_{k,l} = \sigma_0^2\psi_{0,k-l}^{(b)} \quad \text{for } k \geq l,$$

zeros otherwise, $\mathcal{G}_{ac,C}(\theta_0) = 0_{m\times 1}$, and $\mathcal{G}_{ac,D}(\theta_0) = 0_{m\times d}$.

Matrix $\Upsilon_{ac}(\theta_0)$ is a 2×2 block matrix in (21). The upper left matrix was already introduced above. The lower right matrix $\Upsilon_{ac,\mathcal{H}}(\theta_0)$ is an $m \times m$ diagonal matrix

$$\Upsilon_{ac,\mathcal{H}}(\theta_0) = \sigma_0^4 I_{m\times m}.$$

The off-diagonal blocs $\Upsilon_{ac,\Psi}(\theta_0)$ can be written as a 1×4 block matrix

$$\Upsilon_{ac,\Psi}(\theta_0) = \left[\underset{m\times P}{\Upsilon_{ac,\Psi,A}(\theta_0)} \quad \underset{m\times Q}{\Upsilon_{ac,\Psi,B}(\theta_0)} \quad \underset{m\times 1}{\Upsilon_{ac,\Psi,C}(\theta_0)} \quad \underset{m\times d}{\Upsilon_{ac,\Psi,D}(\theta_0)}\right],$$

where the blocks have the typical (k,l) elements,

$$\begin{aligned} [\Upsilon_{ac,\Psi,A}(\theta_0)]_{k,l} &= \sigma_0^2 \psi_{0,k-l}^{(a)} \quad \text{for } k \geq l, \\ [\Upsilon_{ac,\Psi,B}(\theta_0)]_{k,l} &= -\sigma_0^2 \psi_{0,k-l}^{(b)} \quad \text{for } k \geq l, \end{aligned}$$

zeros otherwise, and $\Upsilon_{ac,\Psi,C}(\theta_0) = 0_{m\times 1}$, and $\Upsilon_{ac,\Psi,D}(\theta_0) = 0_{m\times d}$.

Using the fact that $\mathcal{G}_{ac}(\theta_0) = -\Upsilon_{ac,\Psi}(\theta_0)$ and the properties in Proposition 1, we can further simplify the asymptotic covariance matrix as

$$\Omega_{ac} = \mathcal{G}_{ac}(\theta_0)\mathcal{J}(\theta_0)^{-1}\mathcal{G}_{ac}(\theta_0)' + \Upsilon_{ac,\mathcal{H}}(\theta_0).$$

Asymptotic covariance matrix Ω_{hs}

For the $Q_{hs,T}$ test, the matrix $\mathcal{G}_{hs}(\theta_0) = 0_{m\times(P+Q+1+d)}$. This simplifies the covariance as $\Omega_{hs} = \Upsilon_{hs,\mathcal{H}}(\theta_0)$. This matrix, in turn, is a diagonal matrix

$$\Omega_{hs} = \mathrm{E}\left[(\varepsilon_t^2 - \sigma_0^2)^2\right]^2 I_{m\times m}.$$

Consistent estimators for Ω_{ac} and Ω_{hs}

A consistent estimator for the $\mathcal{J}(\theta_0)$ is given by the (feasible) Hessian of the log-likelihood function $\tilde{L}_{\theta\theta,T}(\tilde{\theta}_T) \overset{a.s.}{\to} \mathcal{J}(\theta_0)$ as $T \to \infty$ (this follows from Proposition 1). Note, however, that in the noninvertible case, the outer product of the score does not converge to the same limit (Meitz and Saikkonen, 2013). In order to estimate the rest of the Ω_{ac} consistently, we suggest replacing the parameters appearing in $\mathcal{G}_{ac}(\theta_0)$ and $\Upsilon_{ac,\mathcal{H}}(\theta_0)$ by their estimated values. Almost sure convergence follows from the continuity of $\mathcal{G}_{ac}(\theta)$ and $\Upsilon_{ac,\mathcal{H}}$ and the continuous mapping theorem.

We estimate the $\Omega_{hs,T}$ by replacing the diagonal elements by their consistent estimator $T^{-1}\sum_{t=1}^{T}\left(\tilde{u}_t(\tilde{\theta}_T)^2-\tilde{\sigma}_T^2\right)^2$. Consistency of this estimator is justified by the a.s. consistency $\tilde{\theta}_T \to \theta_0$, continuity of $u_t(\theta)$ and $\tilde{u}_t(\theta)$ for all $\theta \in \Theta_0$, and following two lemmas,

Lemma E4. *Under Assumptions 1 (b), 2 and 3,*

$$\sup_{\theta\in\Theta_0}\left|T^{-1}\sum_{t=1}^{T}\left(u_t(\theta)^2-\sigma^2\right)^2-T^{-1}\sum_{t=1}^{T}\left(\tilde{u}_t(\theta)^2-\sigma^2\right)^2\right|\to 0 \ a.s.\ as\ T\to\infty.$$

Lemma E5. *Under Assumptions 1 (b), 2 and 3,*

$$\sup_{\theta\in\Theta_0}\left|T^{-1}\sum_{t=1}^{T}\left(u_t(\theta)^2-\sigma^2\right)^2-E\left[\left(u_t(\theta)^2-\sigma^2\right)^2\right]\right|\to 0 \ a.s,\ as\ T\to\infty.$$

Proof. Proofs of Lemmas E4 and E5 are given in the supplementary appendix. □

Under assumptions in Theorem 1 we have

$$T^{-1}\sum_{t=1}^{T}\left(\tilde{u}_t(\tilde{\theta}_T)^2-\tilde{\sigma}_T^2\right)^2 I_{m\times m}\to\Omega_{hs}\ a.s.\ as\ T\to\infty.$$

As pointed out in Remark 1, using this estimator for $\Omega_{hs,T}$, the $Q_{hs,T}$ test statistic coincides numerically with the McLeod-Li portmanteau Q-statistic. An alternative but asymptotically equivalent test statistic can be found by making use of the fourth moments of ε_t and the employed distributional assumption. For example, the raw moments of Student's t-distribution are defined via the degrees of freedom parameter λ, for which we have a consistent estimator $\tilde{\lambda}_T$. This approach is not pursued here any further.